Effect of competing mortality risks on predictive performance of the QRISK3 cardiovascular risk prediction tool in older people and those with comorbidity

Livingstone, Shona; Morales, Daniel R.; Donnan, Peter T.; Payne, Katherine; Thompson, Alexander J.; Youn, Ji-Hee

DOI:
10.1016/S2666-7568(21)00088-X

Publication date:
2021

Effect of competing mortality risks on predictive performance of the QRISK3 cardiovascular risk prediction tool in older people and those with comorbidity: external validation population cohort study

Shona Livingstone, Daniel R Morales, Peter T Donnan, Katherine Payne, Alexander J Thompson, Ji-Hee Youn, Bruce Guthrie

Lancet Healthy Longev 2021; 2: e352–61

See **Comment** page e308

Division of Population Health and Genomics, University of Dundee, Dundee, UK (S Livingstone MSc, D R Morales PhD, Prof P T Donnan PhD); **Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, UK** (Prof K Payne PhD, A J Thompson PhD, J-H Youn PhD); **Usher Institute, University of Edinburgh, Edinburgh, UK** (Prof B Guthrie PhD)

Correspondence to: Prof Bruce Guthrie, Old Medical School, University of Edinburgh, Edinburgh EH8 9AG, UK **bruce.guthrie@ed.ac.uk**

Summary

Background Primary prevention of cardiovascular disease (CVD) is guided by risk-prediction tools, but these rarely account for the risk of dying from other conditions (ie, competing mortality risk). In England and Wales, the recommended risk-prediction tool is QRISK2, and a new version (QRISK3) has been derived and internally validated. We aimed to externally validate QRISK3 and to assess the effects of competing mortality risk on its predictive performance.

Methods For this retrospective population cohort study, we used data from the Clinical Practice Research Datalink. We included patients aged 25–84 years with no previous history of CVD or statin treatment who were permanently registered with a primary care practice, had up-to-standard data for at least 1 year, and had linkage to Hospital Episode Statistics discharge and Office of National Statistics mortality data. We compared the QRISK3-predicted 10-year CVD risk with the observed 10-year risk in the whole population and in important subgroups of age and multimorbidity. QRISK3 discrimination and calibration were examined with and without accounting for competing risks.

Findings Our study population included 1 484 597 women with 42 451 incident CVD events (4·9 cases per 1000 person-years of follow-up, 95% CI 4·89–4·99), and 1 420 176 men with 53 066 incident CVD events (6·7 cases per 1000 person-years, 6·66–6·78), with median follow-up of 5·0 years (IQR 1·9–9·2). Non-CVD death rose markedly with age (0·4% of women and 0·5% of men aged 25–44 years had a non-CVD death *vs* 20·1% of women and 19·6% of men aged 75–84 years). QRISK3 discrimination in the whole population was excellent (Harrell's C-statistic 0·865 in women and 0·834 in men) but was poor in older age groups (<0·65 in all subgroups aged 65 years or older). Ignoring competing risks, QRISK3 calibration in the whole population and in younger people was excellent, but there was significant over-prediction in older people. Accounting for competing risks, QRISK3 systematically over-predicted CVD risk, particularly in older people and in those with high multimorbidity.

Interpretation QRISK3 performed well at the whole population level when ignoring competing mortality risk. The tool performed considerably less well in important subgroups, including older people and people with multimorbidity, and less well again after accounting for competing mortality risk.

Funding National Institute for Health Research.

Introduction

Cardiovascular disease (CVD) remains a major cause of morbidity and mortality worldwide despite falling incidences in most high-income countries. Guidelines for the primary prevention of CVD usually recommend the use of risk-prediction tools to target treatment for people above a specified threshold of predicted risk. There has been a progressive reduction in the risk threshold recommended in relation to statin prescription for primary prevention. In England and Wales, the recommended threshold for treatment changed from a 10-year CVD risk of 20% to 10% in 2014,[1] compared with a 7·5% threshold in current US guidelines.[2] These reductions reflect both increasing evidence of statin effectiveness for primary prevention and falling prices, making statins more cost-effective at lower levels of baseline risk. However, age is the most important predictor of CVD risk, and thus most people will exceed current thresholds at some point in early older age w years) irrespective of other risk factors.

Risk-stratified guideline recommendations rely on being able to accurately predict the risk of CVD events. Recommended risk-prediction tools differ between countries, reflecting variations in CVD risk factors and incidence. In England and Wales, the National Institute for Health and Care Excellence recommends the QRISK2 risk-prediction tool,[1] which has been externally validated in UK primary care datasets and found to have

Research in context

Evidence before this study

Guidelines for the primary prevention of cardiovascular disease (CVD) usually recommend risk-stratified treatment. Decisions to start long-term medication to prevent future CVD events are guided by estimation of CVD risk, with treatment offered if patients exceed a particular risk threshold. Recommended risk-prediction tools vary by country, reflecting differences in CVD risk factors and incidence. The recommended risk-prediction tool in the UK is QRISK, but there are two criticisms of recommended tools: first, they often do not predict risk well in older people and people with multimorbidity and second, they do not account for competing mortality risk (the risk of dying from non-CVD causes). We searched PubMed from inception to Jan 8, 2021, for observational studies in English examining competing mortality risks in people with CVD or in the context of incident CVD risk prediction using the search terms (cardiovascular disease[MeSH Major Topic] AND "competing risk" AND ("heart disease risk factors"[MESH Terms] OR prediction)). We found 12 relevant studies examining over-estimation of CVD rates during follow-up, in the context of incident CVD in the whole population and in high-risk populations, such as people with atrial fibrillation, and in the context of additional CVD-related events in people with established CVD. The degree of over-estimation of CVD varied with the population and is believed to be higher in older people for whom competing mortality risk is higher but is not usually accounted for by CVD risk-prediction tools.

Added value of this study

This study shows that CVD risk is systematically over-predicted in older people and in those with more long-term conditions once the competing risk of non-CVD death is accounted for. These findings add to the evidence that risk prediction of single conditions that does not account for competing risks is unreliable in important subgroups.

Implications of all the available evidence

CVD risk-prediction models need to be validated in older people and in people with high multimorbidity. Better CVD risk-prediction models are needed to stratify people who are potentially eligible for primary preventive treatments. Clinicians should consider competing mortality risks and non-CVD life expectancy when discussing statin initiation for primary prevention in older people and in people with high multimorbidity.

excellent discrimination and calibration at the whole population level.[3] QRISK3, a new version of the QRISK tool, which includes additional morbidities for prediction, has been derived and internally validated with the same methods and UK primary care dataset.[4] In the QRISK3 internal validation, overall model discrimination was excellent, although somewhat lower in older people, and calibration was excellent in younger people and very good in older people,[4] but external validation is required before recommending any prediction tool for routine use.[5–7]

However, there are additional concerns about risk prediction that are not directly addressed by conventional external validation. In particular, people who are more likely to die from non-CVD conditions might have little potential benefit from statins but at least some risk of harm from treatment.[8] The issue is one of competing risk, which in this context arises when an individual is at risk of dying from conditions other than CVD. These are obvious at the extreme—taking a statin is clearly futile in someone at the very end of life. However, even smaller levels of competing risk can lead to systematic over-prediction of CVD risk in people at higher risk of dying from another cause, including older people and those with multimorbidity.[9,10] This is because survival analyses, in which data are censored, usually assume that those lost to follow-up have the same risk of the outcome as those who remain in follow-up (eg, if using the Kaplan-Meier estimator). This assumption is incorrect if someone dies of another condition (competing mortality) because a dead person cannot have a CVD event.[11] In this study, we aimed to externally validate QRISK3 and to examine the effect of competing risk on predictive performance.

Methods

Data source and population

For this population cohort study, we externally validated QRISK3 in the UK Clinical Practice Research Datalink (CPRD) Gold,[12,13] which does not overlap with the derivation dataset, although it is similar in its inclusion of linked primary care, hospital, and mortality data. Included patients were permanently registered with a primary care practice, contributed up-to-standard data for at least 1 year, and had linkage to Hospital Episode Statistics (HES) discharge data and Office of National Statistics (ONS) mortality data; were aged 25–84 years with no previous history of CVD; and had no history of previous statin treatment. Cohort entry was the latest date of Jan 1, 2004, a patient's 25th birthday, or contribution of up-to-standard data for at least 1 year. Cohort exit was the date of a first CVD event, death, prescription of a statin (since the main use of the prediction model is to make decisions about statin initiation), deregistration from the primary care practice, date of the last data collection from the practice, or the end of the study on March 31, 2016, whichever came first. All outcomes and predictors were recorded during routine clinical care and were therefore recorded blind to the study hypothesis. This study was approved by the CPRD Independent Scientific Advisory Committee (protocol 16_248).

Outcomes

A first CVD event was defined as the earliest recording of any fatal or non-fatal coronary heart disease, ischaemic stroke, or transient ischaemic attack. Fatal CVD events were identified with codes from the International Classification of Diseases, tenth version (ICD-10), recorded in ONS death registration. Non-fatal events were identified either in primary care electronic health records (using Read codes, the standard coding system used in UK clinics) or HES discharge diagnoses (ICD-10 codes). Read and ICD-10 codes defining outcomes were those used in QRISK3 derivation (appendix p 2).[4]

Prediction model

We implemented the published QRISK3-2017 prediction model (under GNU Lesser General Public Licence, version 3) with some differences: we chose a later cohort entry date (Jan 1, 2004, rather than Jan 1, 1998), we handled missing cholesterol values differently (if no values were available at baseline, QRISK3 derivation allowed cholesterol values from after the index date to be used if they were measured before any event; instead, we only included values recorded before the index date to avoid using future information in prediction), and we evaluated the Townsend deprivation score as the median of the vigintile (equal 20th) of the score for the area within which an individual lived, as individual values were not available. Predictor code sets used in this study and methods of data handling are detailed in the appendix (pp 3–12, 21–165).

Multimorbidity

For each patient at baseline, we additionally calculated a modified Charlson Comorbidity Index (mCCI) based on primary care Read codes (modified such that CVD could not contribute to the score, as all participants were CVD-free at baseline). The mCCI was not used for prediction but was used to stratify the population to examine discrimination and calibration by mCCI score (grouped into 0, 1, 2, and 3 or more).

Missing data

The extent and management of missing data is detailed in the appendix (p 14). As was done in QRISK3 derivation, patients with missing Townsend deprivation scores were excluded from the cohort, those with missing data on ethnicity were assumed to be White, and multiple imputation was used for missing body-mass index, total cholesterol to HDL cholesterol ratio, systolic blood pressure and its variability, and smoking status. Multivariate imputation by chained equations[14] was used to generate five imputed datasets. We combined analyses of these imputed datasets using Rubin's rules[15] to give summary point estimates with confidence limits that reflected the added uncertainty associated with imputing missing values. As with QRISK3 derivation, morbidities and prescribing used for prediction were assumed to be absent if not recorded (morbidity and prescribing recording in CPRD is generally good).[12,13]

Statistical analysis

The study size was determined by the data available in CPRD, which was considered sufficient,[5] and no formal power calculation was done.[5]

We calculated the 10-year risk of having a cardiovascular event for each patient using the published QRISK3 equation without recalibration. The performance of the risk score was assessed by examining discrimination and calibration in the whole population and in subgroups defined by age group (25–44, 45–64, 65–74, and 75–84 years) and by mCCI score.

See Online for appendix

Discrimination is the ability of the risk score to differentiate between patients who had the event of interest during the study and those who did not. We used the truncated version of Harrell's C-statistic to only include pairs for which the earliest survival time was no later than 10 years after entry. Where considerable censoring occurs, Harrell's C-statistic might overestimate discrimination. Therefore, we did a sensitivity analysis using a weighted C-statistic accounting for the probability of censoring.[16] A C-statistic of 0·5 indicates discrimination that is no better than chance, whereas a C-statistic of 1 indicates perfect discrimination. Two additional measures of discrimination were calculated: Royston's D-index (based on the prognostic separation in event-free survival between patients with predicted risk scores higher and lower than the median; higher values of Royston's D-index indicate greater discrimination)[17] and a related R^2 statistic estimating the explained variation in the context of censored survival data.[18]

Calibration refers to how closely the predicted and observed probabilities are similar at group level. This was assessed for equally sized groups of participants ranked by predicted risk. Calibration of the risk score predictions was assessed by plotting observed risk of CVD events against predicted risk. Plots were generated separately by sex, for all patients and for prespecified subgroups of age and mCCI, on the basis of summary statistics pooled across the imputed datasets.

The following summary statistics and their SEs were obtained by decile of predicted risk score and for each imputed dataset in turn: non-parametric measures of observed risk or proportions of patients with a CVD event, the Kaplan-Meier estimator (the conventional measure ignoring competing risks) and the Aalen-Johansen estimator (an extension to allow for competing events, non-CVD death in this case),[19] and the mean predicted risk score. All models were fitted in R, version 4.0.0, and STATA, version 11.2.

Role of the funding source

The study funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

	Women		Men	
	External validation cohort (n=1 484 597)	Original QRISK3 internal validation cohort (n=1 360 457)	External validation cohort (n=1 420 176)	Original QRISK3 internal validation cohort (n=1 310 841)
Age, years	46·0 (15·3)	43·3 (15·3)	44·8 (13·9)	42·6 (13·8)
Body-mass index	25·9 (5·7)	25·4 (5·1)	26·6 (4·7)	25·9 (4·2)
Total cholesterol to HDL cholesterol ratio	3·7 (1·1)	3·6 (1·2)	4·4 (1·3)	4·4 (1·3)
Systolic blood pressure, mm Hg	125·4 (18·0)	123·1 (18·1)	131·1 (16·2)	128·8 (16·2)
Systolic blood pressure variability	10·0 (5·7)	9·3 (6·1)	10·3 (6·2)	9·9 (6·8)
Ethnicity				
White or not recorded	1 363 146 (91·8%)	1 218 391 (89·6%)	1 336 221 (94·1%)	1 171 281 (89·4%)
Indian	22 488 (1·5%)	23 146 (1·7%)	15 322 (1·1%)	26 479 (2·0%)
Pakistani	9550 (0·6%)	10 919 (0·8%)	6647 (0·5%)	14 787 (1·1%)
Bangladeshi	2594 (0·2%)	8738 (0·6%)	2145 (0·2%)	11 914 (0·9%)
Other Asian	13 697 (0·9%)	17 078 (1·3%)	9973 (0·7%)	15 966 (1·2%)
Black Caribbean	9505 (0·6%)	13 142 (1·0%)	6687 (0·5%)	10 642 (0·8%)
Black African	18 804 (1·3%)	27 678 (2·0%)	12 822 (0·9%)	25 251 (1·9%)
Chinese	6739 (0·5%)	8992 (0·7%)	3503 (0·2%)	6098 (0·5%)
Other	38 074 (2·6%)	32 373 (2·4%)	26 829 (1·9%)	28 423 (2·2%)
Smoking status (% of non-missing)				
Non-smoker	707 774 (59·8%)	706 671 (51·9%)	478 671 (49·0%)	512 252 (39·1%)
Former smoker	217 404 (18·4%)	194 545 (14·3%)	216 883 (22·2%)	196 459 (15·0%)
Light smoker	85 277 (7·2%)	154 565 (11·4%)	75 260 (7·7%)	177 693 (13·6%)
Moderate smoker	111 690 (9·4%)	74 933 (5·5%)	112 411 (11·5%)	84 914 (6·5%)
Heavy smoker	62 236 (5·3%)	38 218 (2·8%)	93 457 (9·6%)	64 107 (4·9%)
Family history of CHD*	97 624 (6·6%)	164 023 (12·1%)	75 237 (5·3%)	123 039 (9·4%)
Type 1 diabetes	3752 (0·3%)	3351 (0·2%)	4843 (0·3%)	3932 (0·3%)
Type 2 diabetes	17 022 (1·1%)	15 872 (1·2%)	21 077 (1·5%)	19 318 (1·5%)
Treated hypertension	115 944 (7·8%)	77 694 (5·7%)	82 768 (5·8%)	56 920 (4·3%)
Rheumatoid arthritis	12 702 (0·9%)	15 139 (1·1%)	4724 (0·3%)	7055 (0·5%)
Atrial fibrillation	8199 (0·6%)	5229 (0·4%)	10 620 (0·7%)	6874 (0·5%)
Chronic kidney disease (stage 3, 4, or 5)	6918 (0·5%)	6949 (0·5%)	5659 (0·4%)	4232 (0·3%)
Migraine	117 692 (7·9%)	89 504 (6·6%)	41 471 (2·9%)	36 141 (2·8%)
Corticosteroid use	20 674 (1·4%)	31 775 (2·3%)	11 824 (0·8%)	18 634 (1·4%)
HIV or AIDS	289 (<0·1%)	1595 (0·1%)	445 (<0·1%)	2945 (0·2%)
Systemic lupus erythematosus	1725 (0·1%)	1349 (0·1%)	165 (<0·1%)	134 (<0·1%)
Atypical antipsychotic use	8469 (0·6%)	6268 (0·5%)	8336 (0·6%)	6597 (0·5%)
Severe mental illness	110 799 (7·5%)	94 724 (7·0%)	57 264 (4·0%)	57 830 (4·4%)
Erectile dysfunction diagnosis or treatment	NA	NA	39 264 (2·8%)	31 136 (2·4%)

Data are mean (SD) or n (%). CHD=coronary heart disease. NA=not applicable. *In a first-degree relative younger than 60 years.

Table 1: **Baseline data in external validation cohort and in original QRISK3 internal validation cohort[8]**

Results

Of the patients assessed aged 25–84 with linkage to HES and ONS, 1650 188 were women, of whom 165 591 (10·0%) were excluded because of data inconsistencies (1405 [0·1%]), previous CVD (78 032 [4·7%]), statin prescription (83 357 [5·1%]), or missing deprivation score (2797 [0·2%]); and 1 623 394 were men, of whom 203 218 (12·5%) were excluded because of data inconsistencies (1815 [0·1%]), previous CVD (112 073 [6·9%]), statin prescription (86 656 [5·3%]), or missing deprivation score (2674 [0·2%]). Therefore, 1 484 597 women and 1 420 176 men were included in this study.

Across most baseline characteristics, the study population and the QRISK3 internal validation cohort[4] were similar (table 1) but, in this study, the prevalence of treated hypertension and current smoking was higher and recorded family history of coronary heart disease was lower. Missing data in this study compared with that in the QRISK3 internal validation cohort was less frequent for ethnicity, similar for systolic blood pressure and body-mass index, and more frequent for total cholesterol to HDL cholesterol ratio, systolic blood pressure variability, and smoking status (appendix p 14).

	Patients entering study cohort	Non-fatal CVD	CVD death	Censored			At least 10 years of follow-up
				Non-CVD death	Started statin	Deregistered or end of study* before 10 years of follow-up	
All women	1 484 597	34 047 (2·3%)	5001 (0·3%)	40 839 (2·8%)	128 183 (8·6%)	926 832 (62·4%)	349 695 (23·6%)
All men	1 420 176	42 675 (3·0%)	6471 (0·5%)	38 226 (2·7%)	145 482 (10·2%)	895 421 (63·1%)	291 901 (20·6%)
Women, age (years)							
25–44	813 157	3064 (0·4%)	124 (<0·1%)	3250 (0·4%)	14 076 (1·7%)	612 336 (75·3%)	180 307 (22·2%)
45–64	465 484	10 825 (2·3%)	671 (0·1%)	11 101 (2·4%)	68 552 (14·7%)	242 367 (52·1%)	131 968 (28·4%)
65–74	121 267	8958 (7·4%)	1142 (0·9%)	9454 (7·8%)	32 139 (26·5%)	43 549 (35·8%)	26 205 (21·6%)
75–84	84 689	11 200 (13·2%)	3064 (3·6%)	17 034 (20·1%)	13 416 (15·8%)	28 760 (34·0%)	11 215 (13·2%)
mCCI							
0	1 187 965	21 890 (1·8%)	2908 (0·2%)	22 287 (1·9%)	86 730 (7·3%)	769 100 (64·7%)	285 050 (24·0%)
1	229 651	7981 (3·5%)	1273 (0·6%)	9272 (4·0%)	28 553 (12·4%)	128 966 (56·2%)	53 606 (23·3%)
2	51 295	2956 (5·8%)	567 (1·1%)	6211 (12·1%)	9787 (19·1%)	22 698 (44·2%)	9076 (17·7%)
≥3	15 686	1220 (7·8%)	253 (1·6%)	3069 (19·6%)	3113 (19·8%)	6068 (38·7%)	1963 (12·5%)
Men, age (years)							
25–44	815 950	5659 (0·7%)	461 (0·1%)	4205 (0·5%)	25 050 (3·1%)	614 615 (75·3%)	165 960 (20·3%)
45–64	458 384	19 595 (4·3%)	2105 (0·5%)	12 211 (2·7%)	86 437 (18·9%)	234 266 (51·2%)	103 770 (22·6%)
65–74	96 404	9870 (10·2%)	1607 (1·7%)	9572 (9·9%)	26 821 (27·8%)	31 910 (33·1%)	16 624 (17·2%)
75–84	49 438	7551 (15·3%)	2298 (4·6%)	12 238 (24·8%)	7174 (14·5%)	14 630 (29·6%)	5547 (11·2%)
mCCI							
0	1 173 065	30 524 (2·6%)	4269 (0·4%)	22 906 (2·0%)	104 942 (8·9%)	763 831 (65·1%)	246 593 (21·0%)
1	201 200	8228 (4·1%)	1368 (0·7%)	7903 (3·9%)	29 919 (14·9%)	113 921 (56·6%)	39 861 (19·8%)
2	34 665	2814 (8·1%)	549 (1·6%)	4758 (13·7%)	8088 (23·3%)	13 994 (40·4%)	4462 (12·9%)
≥3	11 246	1109 (9·9%)	285 (2·5%)	2659 (23·6%)	2533 (22·5%)	3675 (32·7%)	985 (8·8%)

Data are n or n (%). CVD=cardiovascular disease. mCCI=modified Charlson Comorbidity Index. *March 31, 2016.

Table 2: **Follow-up and censoring events at 10 years**

42 451 incident cases of CVD were observed in women during 8 594 620 years of follow-up (4·9 cases per 1000 person-years, 95% CI 4·89–4·99), compared with 53 066 incident cases in men during 7 896 704 years of follow-up (6·7 cases per 1000 person-years, 6·66–6·78). CVD incidence rose progressively with age (appendix p 15) and was moderately lower than that observed in QRISK3 derivation.[4] Median follow-up in the whole cohort was 5·0 years (IQR 1·9–9·2), with 641 596 (22·1%) of 2 904 773 patients remaining in the cohort and CVD event-free at 10-year follow-up. By 10 years, CVD events occurred in 39 048 (2·6%) of 1 484 597 women compared with 49 146 (3·5%) of 1 420 176 men, and non-CVD deaths occurred in 40 839 (2·8%) women compared with 38 226 (2·7%) men (table 2). Censoring due to statin initiation was more common than that due to non-CVD death, but almost two thirds of both men and women were censored due to deregistration or to having less than 10 years of follow-up before the end of the study (table 2). Patterns of censoring were markedly different by age and by mCCI. Censoring due to statin initiation rapidly increased with increasing levels of multimorbidity and with age, peaking in the 65–74 years age group in which approximately a quarter of men and women started statins for primary prevention during follow-up. Censoring due to deregistration or having less than 10 years of follow-up by the end of study was more common in younger, than in older, participants and in participants with lower multimorbidity, whereas censoring due to non-CVD death was more common in older participants and people with higher multimorbidity.

Overall discrimination was excellent and similar to that of QRISK3 internal validation[4] (for women, Harrell's C-statistic 0·865 for external validation *vs* 0·880 for internal validation, D 2·43 *vs* 2·49, R^2 58·5% *vs* 59·6%; for men, Harrell's C-statistic 0·834 *vs* 0·858, D 2·10 *vs* 2·26, R^2 51·3% *vs* 55·0%; table 3). However, discrimination varied markedly within the age group and mCCI categories, with discrimination being best in the youngest group (25–44 years) and the group with lowest multimorbidity (mCCI 0) and worst in the oldest group (75–84 years) and the group with highest multimorbidity (mCCI ≥3). Sensitivity analysis using a censoring-adjusted C-statistic found a somewhat lower discrimination than in the main analysis, but did not alter the overall interpretation (appendix p 16).

Ignoring competing mortality risks, calibration was excellent for women overall, and also excellent for

	Women			Men		
	Harrell's C (95% CI)	Royston's D (95% CI)	R^2 (95% CI)	Harrell's C (95% CI)	Royston's D (95% CI)	R^2 (95% CI)
All patients	0·865 (0·861–0·868)	2·43 (2·41–2·45)	58·5% (58·1–58·8)	0·834 (0·831–0·837)	2·10 (2·08–2·12)	51·3% (50·8–51·7)
Age group, years						
25–44	0·758 (0·747–0·769)	1·69 (1·63–1·76)	40·7% (38·8–42·5)	0·757 (0·749–0·764)	1·57 (1·52–1·61)	36·9% (35·6–38·2)
45–64	0·707 (0·702–0·713)	1·25 (1·22–1·28)	27·2% (26·1–28·3)	0·681 (0·677–0·685)	1·04 (1·02–1·07)	20·6% (19·8–21·4)
65–74	0·641 (0·635–0·647)	0·82 (0·77–0·86)	13·7% (12·4–15·1)	0·612 (0·606–0·617)	0·63 (0·59–0·66)	8·6% (7·7–9·5)
75–84	0·611 (0·605–0·616)	0·61 (0·56–0·66)	8·1% (6·9–9·3)	0·585 (0·579–0·591)	0·46 (0·42–0·51)	4·9% (4·1–5·8)
mCCI						
0	0·863 (0·859–0·867)	2·40 (2·38–2·43)	57·9% (57·4–58·4)	0·827 (0·824–0·831)	2·02 (2·00–2·04)	49·4% (48·9–49·8)
1	0·846 (0·840–0·852)	2·20 (2·17–2·24)	53·6% (52·8–54·4)	0·829 (0·823–0·835)	2·00 (1·96–2·03)	48·7% (47·8–49·6)
2	0·789 (0·778–0·799)	1·73 (1·67–1·78)	41·6% (39·9–43·2)	0·728 (0·717–0·739)	1·28 (1·22–1·34)	28·1% (26·2–29·9)
≥3	0·744 (0·728–0·760)	1·40 (1·32–1·48)	31·8% (29·2–34·4)	0·695 (0·678–0·712)	1·13 (1·04–1·21)	23·2% (20·5–26·0)

mCCI=modified Charlson Comorbidity Index.

Table 3: **Discrimination and model fit**

women aged 25–44 years (figure 1; appendix pp 17–18). However, QRISK3 over-predicted CVD risk in older age groups. When stratifying by mCCI, we found evidence of some over-prediction in the group with lowest multimorbidity (mCCI 0) and poor calibration and under-prediction in patients with highest multimorbidity (mCCI ≥3).

When competing mortality risks were accounted for (figure 1), we observed over-prediction of risk at higher levels of predicted CVD risk in all women. The same pattern of increasing over-prediction with increasing age was observed, but in greater magnitude, and calibration was poor in older age groups. Although we observed some under-prediction of risk in patients with mCCI scores of 3 or higher for those at lower predicted CVD risk, the overall pattern was over-prediction of CVD risk, which increased with multimorbidity, and poor calibration in the highest multimorbidity groups (mCCI 2 and ≥3).

Ignoring competing mortality risks, calibration was excellent for all men, although with somewhat greater over-prediction at higher levels of predicted CVD risk than in women (figure 2; appendix pp 19–20). Calibration was excellent for men aged 25–44 years, but QRISK3 progressively over-predicted CVD risk with increasing age. We observed evidence of some over-prediction in the group with lowest multimorbidity (mCCI 0) and poor calibration and under-prediction in the group with highest multimorbidity (mCCI ≥3).

When competing mortality risks were accounted for (figure 2), we observed over-prediction of risk at higher levels of predicted CVD risk in all men. Calibration was poor, with large over-prediction in older age groups. Although we observed some under-prediction of risk in patients with mCCI scores of 3 or higher for those at lower predicted CVD risk, the overall pattern was over-prediction of CVD risk, which increased with multimorbidity, and poor calibration in the groups with highest multimorbidity (mCCI 2 and ≥3).

Discussion

This external validation study found that, at the whole population level, QRISK3 had excellent discrimination overall (the ability of the model to distinguish people at higher or lower risk. However, as is expected when examining discrimination in subsets of the modelled population,[20] discrimination was moderate at best in older people and in people with high levels of multimorbidity. Calibration (the extent to which predicted and observed event rates are similar) was excellent in the whole population when ignoring competing mortality risks, but we found evidence of systematic over-prediction of CVD risk after competing risks were accounted for. Calibration was considerably worse in older people and in those with higher levels of multimorbidity, for whom QRISK3 systematically over-predicted risk, particularly after competing mortality risks were accounted for.

At the whole population level, QRISK3 does appropriately sort the whole population into groups with varying levels of cardiovascular risk (with some small over-prediction), but the model performs relatively poorly in older people and people with high multimorbidity, partly because of high competing mortality risk.

The strengths of our study include methodological conduct consistent with recommendations,[6,21] comprehensive detailing of all code sets to facilitate replication, and explicit assessment of prediction in subgroups and competing mortality risks.

The limitations of this study largely reflect problems common to all studies using routine primary care data, including the original QRISK3 derivation.[22] The prevalence of missing data for key predictors was high. As was done in the original QRISK3 derivation, we used multiple imputation for missing data but, in this context, the assumption that data are missing at random is a strong one because risk factors are plausibly more likely to be measured in people at higher risk (as observed in other studies).[10] However, this limitation of routine data compared with research

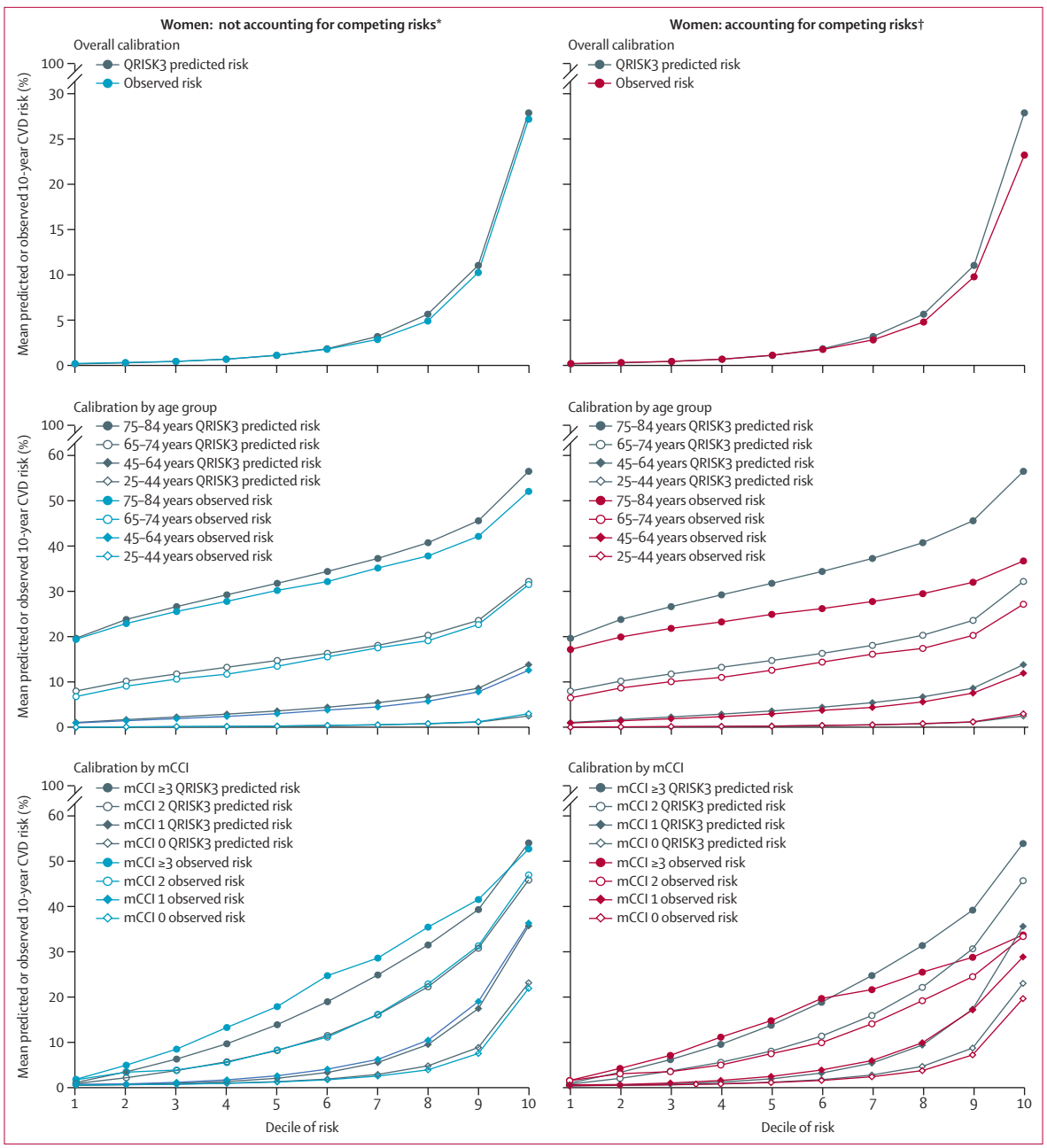

Figure 1: **Calibration in women without accounting for competing risks (left) and accounting for competing risks (right)**
CVD=cardiovascular disease. mCCI=modified Charlson Comorbidity Index. *Observed risk was based on the Kaplan-Meier estimator, which does not account for competing mortality risk. †Observed risk was based on the Aalen-Johansen estimator, which accounts for competing mortality risk.

cohorts with fewer missing data is balanced by routine data cohorts being more representative. All recent QRISK models have also used Jan 1, 1998, as the index date (the earliest that patients can enter the study). Therefore, much observed follow-up in model derivation is historical,[22] and there is a trade-off between using an index date in the distant past (when CVD incidence was higher than it presently is) or a more recent index date (in which case more patients are excluded because of previous statin use). Our choice of a more recent index date might partly explain why QRISK3 was observed to over-predict risk in our validation. Deriving clinical prediction tools on increasingly historical data is probably biased,[22] but using more recent data with greater rates of previous statin initiation might also be biased. There is clearly no optimal resolution to this dilemma. Finally, loss to follow-up before a CVD event was common, which is

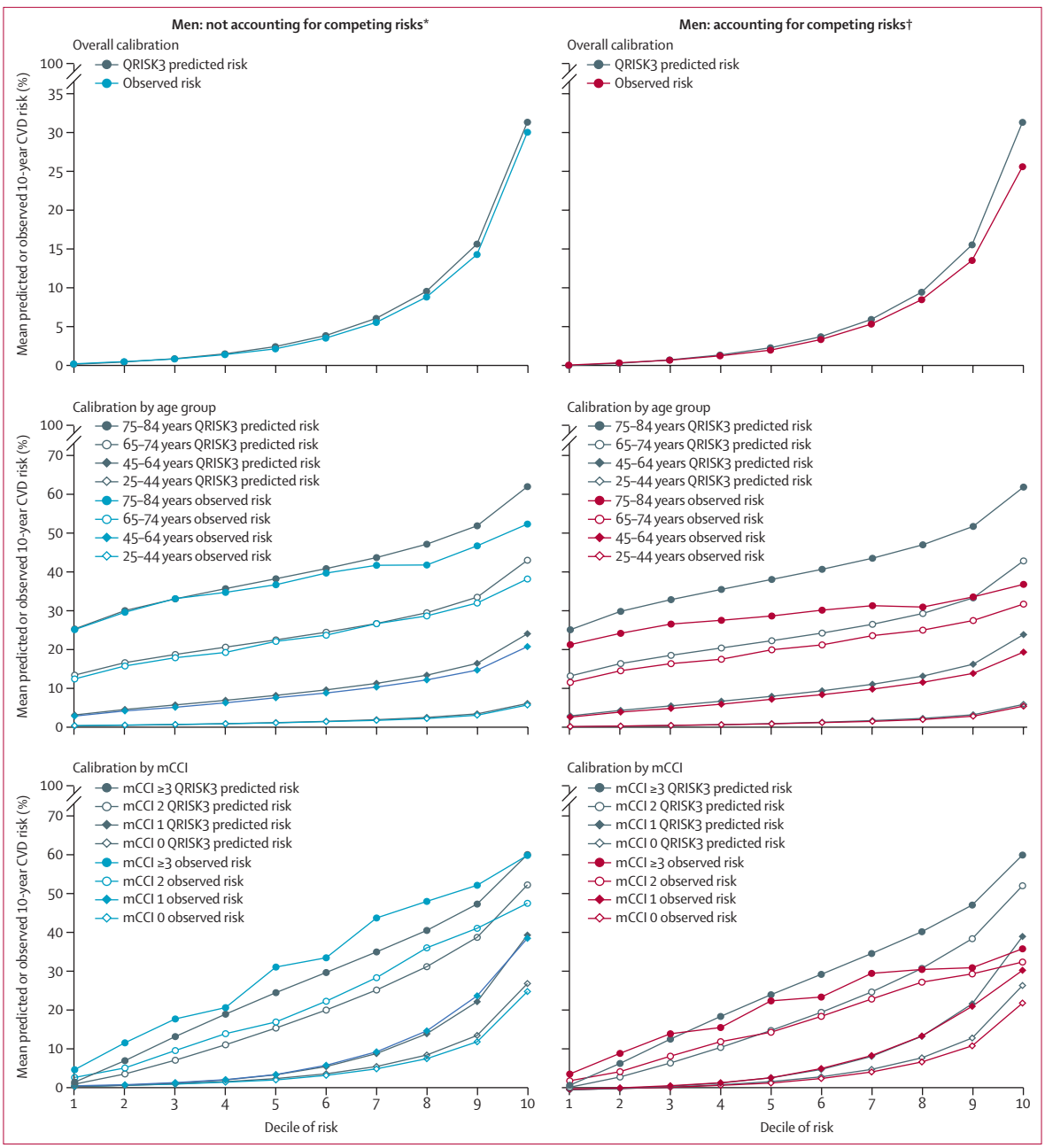

Figure 2: **Calibration in men without accounting for competing risks (left) and accounting for competing risks (right)**
CVD=cardiovascular disease. mCCI=modified Charlson Comorbidity Index. *Observed risk was based on the Kaplan-Meier estimator, which does not account for competing mortality risk. †Observed risk was based on the Aalen-Johansen estimator, which accounts for competing mortality risk.

relevant to model assumptions about censored patients. We specifically examined the effect of censoring due to non-CVD death, but it is also an assumption that those who deregistered from practices had the same CVD risk as those who did not. This seems likely to be the case for younger people, but less so for older people, for whom change of address will be more commonly driven by change in health status (eg, moving to sheltered housing or a care home).

External validations of previous QRISK tools have also found excellent discrimination and calibration at the whole population level when competing mortality risks were ignored (ie, answering the question of what the risk of CVD is, assuming this person does not die of anything else in the following 10 years).[3] Our findings are comparable at the whole population level (ignoring competing risks) but even so, calibration was poor in patients aged 75–84 years and only moderate in those

aged 65–74 years and in those with the highest levels of multimorbidity (mCCI ≥3).

We observed greater over-prediction at the whole population level once competing mortality risk was accounted for. Calibration was notably poorer once competing risk was accounted for, particularly in older patients and those with higher multimorbidity. These findings are consistent with other studies examining the effect of competing risks on estimated CVD risk in people without CVD,[9,11,23,24] with established CVD,[25] and in other contexts, including stroke risk in people with atrial fibrillation.[26,27] QRISK2 has also been shown to systematically over-predict CVD risk in a contemporary population of people with type 2 diabetes, with increasingly poor discrimination with increasing age, highlighting that good performance at the whole population level does not necessarily mean good performance in important subgroups.[20,28]

At the population level, QRISK3 does segment the population into groups in which the observed risk of CVD is very similar to the predicted risk (supporting its use to guide risk-stratified treatment decisions). However, this overall assessment of prediction performance was largely driven by good performance in younger people with fewer coexisting long-term conditions. For older people and people with more long-term conditions, prediction was poor to fair, particularly when competing risks were accounted for. Still, even the lower degree of over-prediction observed in younger and less multimorbid groups can also sometimes change treatment recommendations. Similar issues probably apply to other CVD risk-prediction models that do not account for competing risk. We believe that predicting CVD events without accounting for risk of death from other causes is misleading, particularly in people at high risk of non-CVD death. Therefore, clinicians should carefully consider life expectancy related to other conditions when discussing long-term cardiovascular primary preventive treatment.

Further research would be beneficial in several areas. CVD causes a large proportion of deaths in many high-income countries, which will reduce the effect of competing risks. Additional studies are needed to examine the effect of competing risk when predicting less fatal conditions, for which the effect on predictive performance is likely to be greater. It is also uncertain whether a better approach to CVD prediction would be to create separate models for important subgroups of age and multimorbidity (as is already done for sex), not least because the relationship between classic CVD risk factors and CVD might weaken with age. Further research is needed to evaluate the relative merits of omnibus versus subgroup models, and to better quantify the uncertainty at the individual level of risk-prediction tools that perform well at a population level.[29] A weakness of existing UK primary care datasets in deriving risk-prediction rules is the large loss to follow-up when there is a long time horizon for risk prediction. This study has examined the effect of competing risk, but loss to follow-up due to practice deregistration is likely to create over-prediction in at least some population subsets. External validation in large geographical populations with lower loss to follow-up (such as the SAIL Databank in Wales) would be valuable, as would larger-scale data federation to derive and validate new risk-prediction tools for comparison with QRISK3 and other prediction models.[22] New risk prediction tools could also usefully include statin treatment at baseline in prediction, in the way that QRISK3 includes antihypertensive treatment in prediction.[4] Finally, the value of risk-factor treatment in older people with multimorbidity and co-prescribing who are routinely excluded from trials could be usefully clarified with targeted randomised controlled trials.[30]

In conclusion, QRISK3 performs well at the whole population level but systematically over-predicts CVD risk in older people and people with high multimorbidity. Clinicians should consider broader effects on life expectancy when discussing statin initiation for primary prevention in older people and people with high multimorbidity, in whom CVD risk is likely to be over-predicted. Better calibrated prediction models are needed in these groups.

Contributors

The study was conceived of and designed by BG, KP, DRM, PTD, and AJT who obtained the funding. All authors contributed to study design and interpretation. SL, BG, DRM, and PTD led data management and SL led the analysis, supported by BG, DRM, and PTD. SL and BG drafted the paper, which all authors reviewed and edited. SL, BG, and DRM verified the underlying data. The corresponding author had full access to all the data and the final responsibility to submit for publication.

Declaration of interests

PTD reports a grant from AbbVie, outside the submitted work, and is a member of the NHS Scottish Medicines Consortium. J-HY is currently employed by ICON PLC Clinical Research. DRM reports grants from the Chief Scientist Office, Health Data Research UK, and National Institute for Health Research, outside the submitted work. All other authors declare no competing interests.

Data sharing

The data controller is the CPRD and, under the data licence granted, we are not allowed to share data. Researchers can apply to CPRD directly for access to the raw data.

Acknowledgments

This study was funded by the National Institute for Health Research (NIHR) Health Services and Delivery Research Programme (project reference 15/12/22). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The authors had full and sole access to the data, and the funder had no role in the conduct of the research or the decision to publish. DRM is supported by a Wellcome Trust Clinical Research Development Fellowship (Grant 214588/Z/18/Z).

References

1 National Institute for Health and Care Excellence. Clinical Guideline 181: lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. London: National Institute for Health and Care Excellence, 2014.

2 Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *Circulation* 2014; **129** (suppl 2): S1–45.

3 Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. *BMJ* 2012; **344:** e4181.

4 Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017; **357:** j2099.

5 Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer, 2009.

6 Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014; **14:** 40.

7 Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013; **13:** 33.

8 Mehta S, Jackson R, Poppe K, Kerr AJ, Pylypchuk R, Wells S. How do cardiovascular risk prediction equations developed among 30–74 year olds perform in older age groups? A validation study in 125 000 people aged 75–89 years. *J Epidemiol Community Health* 2020; **74:** 527–33.

9 Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology* 2009; **20:** 555–61.

10 van Staa T-P, Gulliford M, Ng ESW, Goldacre B, Smeeth L. Prediction of cardiovascular risk using Framingham, ASSIGN and QRISK2: how well do they predict individual rather than population risk? *PLoS One* 2014; **9:** e106455.

11 Nguyen QD, Odden MC, Peralta CA, Kim DH. Predicting risk of atherosclerotic cardiovascular disease using pooled cohort equations in older adults with frailty, multimorbidity, and competing risks. *J Am Heart Assoc* 2020; **9:** e016003.

12 Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44:** 827–36.

13 Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; **69:** 4–14.

14 van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011; **45:** 67.

15 Rubin D. Multiple imputation for nonresponse in surveys. New York: John Wiley and Sons, 1987.

16 Gerds TA, Kattan MW, Schumacher M, Yu C. Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Stat Med* 2013; **32:** 2173–84.

17 Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004; **23:** 723–48.

18 Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; **338:** b605.

19 Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007; **26:** 2389–430.

20 Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; **115:** 928–35.

21 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015; **350:** g7594.

22 Pylypchuk R, Wells S, Kerr A, et al. Cardiovascular disease risk prediction equations in 400 000 primary care patients in New Zealand: a derivation and validation study. *Lancet* 2018; **391:** 1897–907.

23 Koller MT, Leening MJG, Wolbers M, et al. Development and validation of a coronary risk prediction model for older U.S. and European persons in the Cardiovascular Health Study and the Rotterdam Study. *Ann Intern Med* 2012; **157:** 389–97.

24 Li Y, Sperrin M, Ashcroft DM, van Staa TP. Consistency of variety of machine learning and statistical models in predicting clinical risks of individual patients: longitudinal cohort study using cardiovascular disease as exemplar. *BMJ* 2020; **371:** m3919.

25 Melberg T, Nygård OK, Kuiper KK, Nordrehaug JE. Competing risk analysis of events 10 years after revascularization. *Scand Cardiovasc J* 2010; **44:** 279–88.

26 Ashburner JM, Go AS, Chang Y, et al. Influence of competing risks on estimating the expected benefit of warfarin in individuals with atrial fibrillation not currently taking anticoagulants: the Anticoagulation and Risk Factors in Atrial Fibrillation study. *J Am Geriatr Soc* 2017; **65:** 35–41.

27 Abdel-Qadir H, Fang J, Lee DS, et al. Importance of considering competing risks in time-to-event analyses: application to stroke risk in a retrospective cohort study of elderly patients with atrial fibrillation. *Circ Cardiovasc Qual Outcomes* 2018; **11:** e004580.

28 Read SH, van Diepen M, Colhoun HM, et al. Performance of cardiovascular disease risk scores in people diagnosed with type 2 diabetes: external validation using data from the National Scottish Diabetes Register. *Diabetes Care* 2018; **41:** 2010–18.

29 Li Y, Sperrin M, Belmonte M, Pate A, Ashcroft DM, van Staa TP. Do population-level risk prediction models that use routinely collected health data reliably predict individual risks? *Sci Rep* 2019; **9:** 11222.

30 He J, Morales DR, Guthrie B. Exclusion rates in randomized controlled trials of treatments for physical conditions: a systematic review. *Trials* 2020; **21:** 228.